



Research on de-motion blur image processing based on deep learning[☆]

Yanfen Chang

School of Information Engineering, Ningbo University of Finance & Economics, Ningbo, Zhejiang, China



ARTICLE INFO

Article history:

Received 24 January 2019

Revised 20 February 2019

Accepted 20 February 2019

Available online 25 February 2019

Keywords:

Deep learning

Image processing

Blurred image

ABSTRACT

In recent years, with the rapid development of computer technology and network technology, computer vision has been widely used in various scientific fields. Human motion recognition, as an important branch of computer vision, is essentially to classify human motion information in motion images correctly. It has great significance in intelligent monitoring and security, human-computer interaction, motion analysis and other fields. At present, there are still some problems in human motion recognition methods. Firstly, how to extract and characterize the motion information in images has been one of the difficulties in this field; secondly, with the appearance of kinect and other depth cameras, researchers have provided the depth information of human motion images, and how to effectively use these depth information to achieve human motion recognition and classification is also an important research issue; finally, when the amount of sample data is small, how to use the deep learning network model to achieve a higher human motion recognition rate? Based on UTD-MHAD database, this paper studies the human motion recognition of RGB image and depth image captured simultaneously by kinect, and carries out relevant discussion and analysis on the above problems, using micro-inertial sensors (MTi-G-700 developed by Xsens and Android mobile phones, tablets and other personal mobile devices come with MEMS gyroscopes and accelerometers) to correct the image to motion blur, build a new mathematical model, use the inertial data obtained by MIMU in a short time to estimate the position, attitude and speed of camera motion, correct the image pixel position, perform image de-motion blur processing, and then perform image processing such as denoising to solve the image motion blur problem. A new algorithm is developed and its science is verified by MATLAB simulation.

© 2019 Published by Elsevier Inc.

1. Introduction

Nowadays, with the rapid development of social science and technology [1], various industries are pursuing information digitization, intelligence, pattern recognition and computer vision and other fields are also constantly carrying out technological innovations [2]. Computer vision is to let computers, cameras and other machines receive information like human beings, analyze and process its semantic information [3], and make coping strategies. Human motion recognition is an important research direction in the field of computer vision. With the gradual rise of deep learning,

there are also new solutions. This paper explores and studies this issue [4].

Image processing is one of the basic technologies in computer applications [5]. It consists of two parts: visual representation and classification and sorting. It mainly uses various descriptors [6], features to represent visual sensing information such as image content, semantics and structure, and then finds corresponding items [7] which are the same or similar to the query image in a large number of image databases. In recent years, with the continuous development of machine intelligent industry and Internet + industry, image processing has attracted more and more attention from researchers. It has gradually penetrated into the lowest level in the fields of neural network, computer network and machine learning, and has become the backbone [8] for social service [9] and industrial upgrading.

As one of the components of image processing, the main research work of visual expression of images is to help computer systems acquire and understand image information [10], so as to obtain information analysis and processing capabilities similar to

Abbreviations: MAE, mean absolute error; MSE, mean square error; SNR, signal to noise ratio; PSNR, peak signal to noise ratio; GMG, grayscale mean gradient; LS, Laplacian summation; ANN, artificial neural network; CNN, convolutional neural networks; DNN, deep neural network; RBM, restricted Boltzmann machines; SVM, support vector machine; DBN, deep belief network; LSTM, long short term memory; LBP, local binary patterns; ELM, extreme learning machine; LTP, long time potentiation.

[☆] This article is part of the Special Issue on TIUSM.

E-mail address: cyf511@163.com

human thinking and vision. On the other hand, image classification and sorting, as a process directly affecting the retrieval results, requires that [11]: in the process of classification, the computer should complete the establishment of visual concepts from limited image information, and be used to obtain and identify the category information of images [12]; it is equally important with the analogy sorting process, and requires that the similarity comparison between image features be completed in a short time, and to complete the ranking order of query images related to candidate images according to similarity. The two main parts of image processing are not only the frontier of this research, but also widely used in other fields. They are closely concerned by academia and industry [13]. For example, scholars from all walks of life have launched a large-scale visual recognition contest for the Image Net image data set, and have continuously developed and expanded the Internet search engine companies: Baidu, Google and so on [14].

At present, image processing still remains a time-consuming and laborious method based on text or keywords in most industrial applications [15]. Of course, content-based image retrieval (CBIR) has been widely distributed in more advanced search engines, but often a single descriptor cannot complete the search of large-scale images and the stability is not high, and the retrieval effect of complex background targets is usually unsatisfactory. With the continuous development of economy and society, the daily image and video frame data generated on the network show exponential growth [16]. It is obvious that this retrieval method based on general global features or numerous local features cannot meet the development momentum of productivity in terms of accuracy or time consumed [17]. On the other hand, with the advent of deep learning convolutional neural network (CNN) in the field of target recognition in 2012 [18], a new image expression came into being, which uses the output of one layer in the network as another expression of data [19], so it can be considered as a feature learned through the network. Based on these characteristics, further similarity comparison can be made. Although the advantages of image feature memory occupancy [20] and similarity comparison are obvious, due to the short birth time, the feature performance of deep learning method is still insufficient, resulting in a certain distance of retrieval accuracy compared with content-based image processing method. Therefore, the development of image processing based on these two methods still has a long way to go, making the continuous improvement of accuracy and optimization model become the goal of every company and researchers day and night [21].

The image processing technology based on CNN was born not long ago [22], but it has already been fruitful. This technique is based on well-trained CNN model, and uses image mapping in convolution layer or full connection layer as feature for subsequent retrieval operation [23]. At first, Babenko et al. proposed to extract the vectors of the full connection layer of the image in the network as the global feature of the image [24]. Although it can meet the requirements of retrieval, it is not sensitive to the spatial description of the image content and the significant region of the image. In order to solve this problem, two major schemes have been accepted by scientific researchers. One is to refer to the convolution layer feature which is more close to the original image, and is also proposed by Babenko et al. [25]. Using sum-pooling strategy, the convolution layer features are processed to form the global features of the image. On this basis, Kalantidis et al. developed weighted cross-dimensional weighting for aggregated deep convolutional feature CroW [26]. By weighting the convolutional feature, the salience of the main area of the image is enhanced, thus reducing the complexity of the image. Another method, many researchers focus on the combination of sliding frame and convolution neural network to localize the full connection layer [27], similar

to the multi-scale orderless pooling of deep convolution activation features or Mop [28].

With the development of convolution neural network and image processing, sliding frame technology is applied to convolution layer to realize local regionalization of image features [29]. Among them, the most classical one belongs to the regional max-pooling of CNN activation R-mac proposed by Tolias et al. [30], which uses sliding frames to slide regularly on convolution layer to extract local features [31]. As a result, more and more improved methods continue to emerge [32]. Gordo and others have increased the recognition function of fixed targets under the R-mac framework, increased the training of external reference databases [33], and greatly improved the retrieval accuracy. Although the design of convolution layer and sliding window has met the requirements of most image processing applications [34], there is a certain gap from the traditional retrieval algorithm, and still need to be improved [35].

In this paper, based on UTD-MHAD database, human motion recognition is studied for RGB images and depth images captured simultaneously by kinect. The above problems are discussed and analyzed. The micro-inertial sensors (MTi-G-700 developed by Xsens and Android mobile phones, tablets and other personal mobile devices come with MEMS gyroscopes and accelerometers) are used to correct the image to remove motion blurring [36] and a new mathematical model is established. In a short time, the inertial data obtained by MIMU are used to estimate the position, attitude and speed of camera motion, correct the position of image pixels, remove motion blurring, and denoise image processing to solve the problem of image motion blurring. A new algorithm is designed, and its scientificness is verified by MATLAB simulation.

2. Proposed method

2.1. De-motion blurring algorithm for single frame image

The sharpening processing algorithm for single frame blurred image is to use the blurred image sharpening processing caused by a single motion, that is, to directly estimate the PSF using the information of the single blurred image itself, and then to perform inverse filtering and other related processing. This kind of algorithm is collectively referred to as the single frame image de-motion blurring algorithm. The motion blurring process of monocular camera image is a typical blind convolution restoration process. The block diagram of the process can be shown in Fig. 1.

That is to say, the motion blurred image acquired by monocular camera is extracted for feature, and the PSF parameters are estimated. The image is inversely filtered by the estimated parameters,

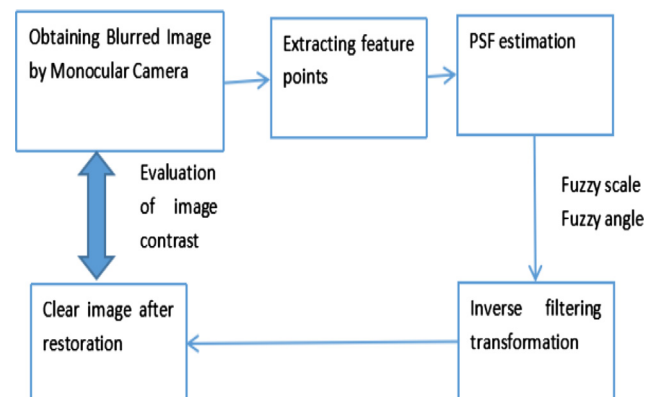


Fig. 1. Block diagram of de-motion blurring process for monocular camera image.

Download English Version:

<https://daneshyari.com/en/article/13436739>

Download Persian Version:

<https://daneshyari.com/article/13436739>

[Daneshyari.com](https://daneshyari.com)